REMARKS

Claims 1-38 remain in this application. Claims 1-3, 5-8, 14-17, 19-22, 28, 30, 31, 34, and 35 have been amended. Claims 1-38 are pending. No new matter has been entered.

Applicant thanks the Examiner for the telephone interview of April 27, 2005, wherein the differences between the cited art and claims were discussed. Claims 1-3, 5-8, 14-17, 19-22, 28, 30, 31, 34, and 35 have been amended to present the rejected claims in better form for consideration on appeal. In addition, the claim amendments were not presented earlier because the Emens reference, cited *supra*, was only first cited in the final Office action to which the present paper responds. Such amendments were necessary to clarify the claimed subject matter in light of the Emens reference. Accordingly, pursuant to 37 C.F.R. 1.116(b), (c), admittance of the claim amendments is respectively requested.

Claims 1-38 stand rejected under 35 U.S.C. § 102(e) as being anticipated by U.S. Patent No. 6,606,643 to Emens et al. ("Emens"). Applicant traverses the rejection. A claim is anticipated under 35 U.S.C. §102(e) only if each and every element as set forth in the claim is found, either expressly or inherently described, in a single prior art reference. MPEP § 2131. The Emens patent fails to teach or suggest each and every claim element and fails to anticipate Claims 1-38.

The Emens patent describes a method of automatically selecting a mirror server from among multiple mirror servers providing Web content for client-host interaction over a network (Abstract). Emens addresses the problem of supplying Web content to a requesting client caused by overloaded servers and network congestion (Col. 1, Lines 33-40). Emens teaches that when a single host server is used to provide Web content, the network connections leading to that server can become a bottleneck, leading to slow response time, even where the server would otherwise be capable of servicing the entire load (Col. 1, Lines 37-40). Emens further teaches that by geographically distributing replicated servers and creating "mirror" sites, such a distributed architecture results in increased availability of the service in times of network congestion and partial unavailability with the potential for increased performance by taking into account topological proximity

of clients and servers (Col. 1, Lines 50-57). Emens further teaches that a current difficulty with distributed mirror server designs is that the user must manually select a server from a list of mirror servers (Col. 1, Lines 58-60).

Thus, Emens teaches a client computer using a Web browser program to send an information request to a central host server, which responds by sending a list of IP addresses in addition to relevant Web content, with each IP address corresponding to an available mirror server (Col. 3, Lines 31-40). The client computer Web browser application initiates multiple concurrently executing threads, each of which corresponds to a unique IP address from the received set of mirror IP addresses, and transparently sends identical mirror server requests to the designated IP addresses (Col. 3, Lines 40-50; Col. 4, Lines 30-54; Col. 9, Lines 26-31). The mirror servers each send a response back to the client computer and each thread measures the roundtrip latency. The thread with the lowest registered roundtrip latency is identified as a "winner," and the Web browser connects with the replicated mirror server site corresponding to the winner thread's associated IP address (Col. 3, Line 54- Col. 4, Line 5; Col. 4, Lines 55-60; Col. 7, Lines 49-54; Col. 8, Lines 25-44).

Consequently, Emens teaches addressing the need of selecting the best mirror server from a number of identical mirror servers. Each thread, called "calibration applets," corresponds to a mirror server on a list of possible mirror servers sent by the host server responsive to the initial information request from the client computer (Col. 7, Lines 39-43). The Web browser application interface interacts directly with the external world and receives the mirror server address list from the host server (Col. 7, Lines 44-47). The calibration applets are responsible for contacting their assigned mirror server and determining roundtrip response time (Col. 7, Lines 48-50). Each calibration applet measures the roundtrip latency average to cross a number of requests and preferably abnormal results identified and discarded (Col. 8, Lines 32-37).

In contrast, independent Claims 1 and 15 respectively define a system and method for efficiently forwarding client requests in a distributed computing environment. Claims 1 and 15 recite receiving a plurality of client requests

commonly routed for forwarding to an origin server from individual sending clients into a proxy server connectively interposed between the sending clients and the origin server. Claims 1 and 15 further recite dynamically generating at the proxy server, concurrent to and during processing of each request by the proxy server, time estimates of service availability based on a time-to-idle for sending the requests over each of a plurality of network connections from the proxy server to the origin server, wherein time-to-idle for each network connection is calculated based on the amount of time that will elapse before an active network connection is usable for a subsequent client request. Finally, Claims 1 and 15 recite selecting the network connection from the proxy server to the origin server with a substantially highest service availability and a substantially lowest time-to-idle and forwarding each request from the proxy server to the origin server using the selected network connection. Support for the claim amendments can be found in the specification on page 6, Line 17- page 7, Line 18; page 8, Line 30- page 9, Line 8; page 10, Lines 10-21. No new matter has been entered.

In contrast, independent Claims 30 and 34 respectively define a system and method for efficiently forwarding requests from a proxy server in a TCP/IP computing environment. Claims 30 and 34 recite receiving a plurality of transient requests from individual sending clients into a request queue on a proxy server connectively interposed between the sending clients and an origin server, each request being commonly routed for forwarding to the origin server. In addition, Claims 30 and 34 recite dynamically calculating at the proxy server, concurrent to receiving and during processing of each request, time estimates of TCP overhead, slow start overhead, time-to-idle, and request transfer time for sending the requests over each of a plurality of managed network connections from the proxy server to the origin server, wherein time-to-idle for each network connection is calculated based on the amount of time that will elapse before an active network connection is usable for a subsequent request. Finally, Claims 30 and 34 recite choosing the managed network connection from the proxy server selected from, in order of preferred selection, a warm idle network connection, an active network connection with a time-to-idle less than a slow start overhead, a cold idle network

connection, an active network connection with a time-to-idle less than a TCP overhead, a new managed network connection, and an existing managed network connection with a smallest time-to-idle; and forwarding each request from the proxy server to the origin server over the selected managed network connection. Support for the claim amendments can be found in the specification on page 6, Line 17- page 7, Line 18; page 8, Line 30- page 9, Line 8; page 10, Lines 10-21. No new matter has been entered.

More particularly, Claims 1, 15, 30, and 34 recite selecting the best connection to a *specific* origin server, which could be on a new network connection or on one of a number of existing network connections, whereas Emens teaches selecting a mirror server from among *multiple* mirror servers (Col. 1, Lines 10-15). Emens teaches away from employing a *single* origin server and recognizes that the network connections leading to a single server can become a bottleneck leading to slow response times (Col. 1, Lines 36-40). Such problem is resolved by the system and method of Claims 1, 15, 30, and 34, which select the best network *connection* to a specific origin server.

Moreover, Claims 1, 15, 30, and 34 recite generating time estimates of service availability based on a time-to-idle for sending requests over each network connection, wherein the time-to-idle is calculated based on the amount of time that will elapse before active network connection is usable for a subsequent client request. Thus, the time estimates of service availability are determined through intimate knowledge of the communications protocol used in the behavior of the origin server. In contrast, the Emens reference teaches performing timing tests by sending multiple, identical mirror server requests and determining the mirror origin server having the best response time based on roundtrip latency averaged over the number of requests (Col. 3, Lines 46-51; Col. 8, Lines 31-37). As a result, Emens teaches assessing which mirror server, out of a number of mirror servers, should be used for communications through a process that can result in wasting network resources through duplicate connections with duplicate client requests being sent to the multiple mirror servers. The amendments to Claims 1, 15, 30, and 34 clarify that client requests commonly routed for forwarding to a

single origin server are received into a proxy server and that the proxy server dynamically generates time estimates of service availability to the origin server over each of a plurality of network connections, wherein time estimates of service availability are based on the amount of time that will elapse before an active network connection is useable for a subsequent client request. Such limitations are neither taught nor suggested by Emens.

Therefore, the Emens reference fails to describe all the claim limitations and does not anticipate Claims 1, 15, 30, and 34. Claims 2-14 are dependent on Claim 1 and are patentable for the above-stated reasons, and as further distinguished by the limitations recited therein. Claims 16-29 are dependent on Claim 15 and are patentable for the above-stated reasons, and as further distinguished by the limitations recited therein. Claims 31-33 are dependent on Claim 30 and are patentable for the above-stated reasons, and as further distinguished by the limitations recited therein. Claims 35-38 are dependent on Claim 34 and are patentable for the above-stated reasons, and as further distinguished by the limitations recited herein. Withdrawal of the rejection under 35 U.S.C. §102(e) is respectfully requested.

The prior art made of record and not relied upon has been reviewed by the applicant and is considered to be no more pertinent than the prior art references already applied.

Claims 1-38 are believed to be in a condition for allowance. Entry of the claim amendments and withdrawal of the finality of the Office action are requested and a Notice of Allowance is earnestly solicited. Please contact the undersigned at (206) 381-3900 regarding any questions or concerns associated with the present matter.

Respectfully submitted,

Dated: May 9, 2005

By: _____
Patrick J.S. Inouye, Esq.
Reg. No. 40,297

Law Offices of Patrick J.S. Inouye

810 Third Avenue, Suite 258
Seattle, WA 98104

Telephone: (206) 381-3900
Facsimile: (206) 381-3999

5 Final OA Response